Zoltán Meggyesi
KFKI - RMKI
Research Institute for Particle and Nuclear Physics
H-1525 Budapest, POB 49, Hungary
E-mail: zoltan.meggyesi@cern.ch

Fibre Channel Overview

Table of Contents

1 Introduction

In recent years several technical developments have converged to a bigger than ever need for extremely fast data links. High performance computers have become the focus of much attention in the data communications industry. Performance improvements have spawned increasingly data-intensive and high-speed networking applications, such as multimedia and scientific visualization. However, the existing network interconnects between computers and I/O devices are unable to run at the speeds needed.

The intention of the Fibre Channel (FC) is to develop practical, inexpensive, yet expendable means of quickly transferring data between workstations, mainframes, supercomputers, desktop computers, storage devices, displays and other peripherials. Fibre Channel is the general name of an integrated set of standards [1] being developed by the American National Standards Institute (ANSI).

There are two basic types of data communication between processors and between processors and peripherials: channels and networks. A channel provides a direct or switched point-to-point connection between the communicating devices. A channel is tipically hardware-intensive and transports data at the high speed with low overhead. In contrast, a Network is an aggregation of distributed nodes (like workstations, file servers or peripherials) with it's own protocol that supports interaction among these nodes. A network has relatively high overhead since it is software-intensive, and consequently slower

than a channel. Networks can handle a more extensive range of tasks than channels as they operate in an environment of unanticipated connections, while channels operate amongst only a few devices with predefined addresses. Fibre Channel attempts to combine the best of these two methods of communication into a new I/O interface that meets the needs of channel users and also network users.

Although it is called Fibre Channel, it's architecture doesn't represent neither a channel nor a real network topology. It allows for an active intelligent interconnection scheme, called a Fabric, to connect devices. All a Fibre channel port has to do is to manage a simple point-to-point connection between itself and the Fabric.

Fibre channel is a high performance serial link supporting its own, as well as higher level protocols such as the FDDI, SCSI, HIPPI and IPI (see chapter 7). The Fibre Channel standard addresses the need for very fast transfers of large amounts of information. The fast (up to 1 Gbit/s) technology can be converted for Local Area Network technology by adding a switch specified in the Fibre Channel standard, that handles multipoint addressing. There is a perspective as an I/O technology and a Local Area Network technology as well. Another advantage of Fibre Channel is, that it gives users one port that supports both channel and network interfaces, unburdening the computers from large number of I/O ports. FC provides control and complete error checking over the link [2] [3].

2 Fibre Channel topology

In Fibre Channel terms the switch connecting the devices is called Fabric. The link is the two unidirectional fibres transmitting to opposite directions with their associated transmitter and receiver. Each fibre is attached to a transmitter of a port at one end and a receiver of another port at the other end. When a Fabric is present in the configuration, the fibre may attach to a node port (N_Port) and to a port of the Fabric (F_Port).

Since Fibre channel system relies on ports logging in with each other and the Fabric, it is irrelevant whether the Fabric is a circuit switch, an active hub or a loop. The topology can be selected depending on system performance requirements or packaging options. Possible FC topologies include point-to-point, crosspoint switched or arbitrated loop (Figure 1).
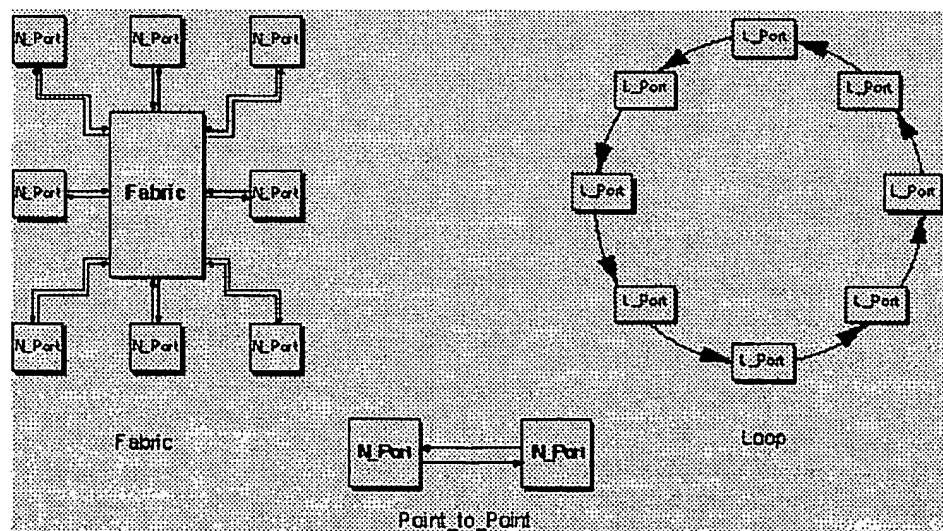
Figure 1 Fibre Channel topologies

FC operates at a wide variety of speeds (133 Mbit/s, 266 Mbit/s, 530 Mbit/s, and 1 Gbits/s) and on three types of both electrical and optical media. Transmission distances vary depending on the combination of speed and media. The single mode fibre optic media using longwave laser light source gives the highest performance (10 km maximum distance at 1 Gbit/s) [2].

Back to Table of Contents

3 FC-0 layer

FC is structured as a set of hierarchical functions (Figure 2).
The lowest level (FC-0) defines the physical link in the system, including the fibre, connectors, optical and electrical parameters for a variety of data rates. Figure 3 shows the schematic of the Fibre Channel optical link [2].
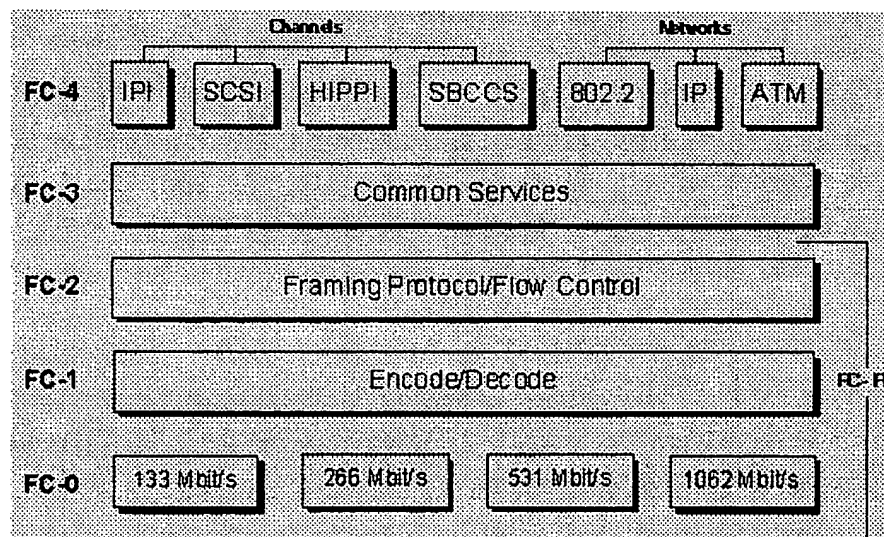

Figure 2 Fibre Channel structure

The system bit error rate (BER) at the supported media and speeds is less than 10exp-12 [1]. The physical level is designed for the use of large number of technologies to meet the widest range of system requirements. An end-to-end communicating route may consist of different link technologies to achieve the maximal performance and price efficiency.

3.1 Open Fibre Control

The FC-0 specifies a safety system - the Open Fibre Control system (OFC) - for SW laser data links, since the optical power levels exceed the limits defined by the laser safety standards. If an open fibre condition occurs in the link, the receiver of the Port the fibre is connected detects it and pulses its laser at a low duty cycle that meets the safety requirements. The receiver of the other port (at the other end of the fibre) detects this pulsing signal and also pulses it's transmitter at a low duty cycle. When the open fibre path is restored both ports receive the pulsing signals, and after a double handshaking procedure the connection is automatically restored within a few seconds [1].
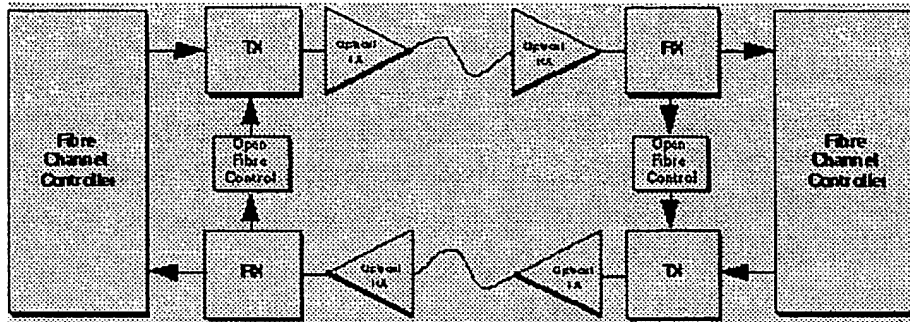
Figure 3 FC optical link

4 FC-1 layer

FC-1 defines the transmission protocol including serial encoding and decoding rules, special characters and error control. The information transmitted over a fibre is encoded 8 bits at a time into a 10 bit Transmission Character. The primary rationale for use of a transmission code is to improve the transmission characteristic of information across a fibre. The transmission code must be DC balanced to support the electrical requirements of the receiving units. The Transmission Characters ensure, that short run lengths and enough transitions are present in the serial bit stream to make clock recovery possible [1] [2].

4.1 FC-1 character conversion

An unencoded information byte is composed of eight information bits A,B,C,D,E,F,G,H and the control variable Z. This information is encoded by FC-1 into the bits a,b,c,d,e,i,f,g,h,j of a 10-bit Transmission Character. The control variable has either the value D (D-type) for Data characters or the value K (K-type) for special characters. Each valid Transmission Character has been given a name using the following convention: Zxx.y, where Z is the control variable of the unencoded FC-1 information byte, xx is the decimal value of the binary number composed of the bits E, D, C, B, and A, and y is the decimal value of the binary number composed of the bits H,G of the unencoded FC-1 information byte in that order. For example the name of the FC-1 Transmission Character composed of the hexadecimal "BC" special (K-type) code is K28.5.

The information received is recovered 10 bits at a time and those Transmission Characters used for data (D-type) are decoded into the one of the 256 8-bit combinations. Some of the remaining Transmission Characters (K-type) referred to as special characters, are used for protocol management functions. Codes detected at the receiver that are not D- or K- type are signaled as code violation errors [1].

4.2 Coding rules

Each data byte or special character has two (not necessarily different) transmission codes. The data bytes and special characters are encoded into these codes respectively, depending on the initial Running Disparity (RD). The RD is a binary parameter, which is calculated upon the balance of ones and zeros in the sub-blocks (the first six bits and the last four bits) of a transmission character. A new RD is calculated from the transmitted character at both the transmitter and the receiver. If the detected character has opposite RD the transmitter should have sent, (depending on the RD of the previous bit

stream) the receiver indicates a disparity violation condition. A Transmission Word is composed of four contiguous transmission characters [1].

5 FC-2 Layer

The Signaling Protocol (FC-2) level serves as the transport mechanism of Fibre Channel. The framing rules of the data to be transferred between ports, the different mechanisms for controlling the three service classes (see chapter 5.7) and the means of managing the sequence of a data transfer are defined by FC-2. To aid in the transport of data across the link, the following building blocks are defined by the standard [1] :

- Ordered Set
- Frame
- Sequence
- Exchange
- Protocol

5.1 Ordered Set

The Ordered Sets are four byte transmission words containing data and special characters which have a special meaning. Ordered Sets provide the availability to obtain bit and word synchronization, which also establishes word boundary alignment. An Ordered Set always begins with the special character K28.5. Three major types of Ordered Sets are defined by the signaling protocol.

The Frame delimiters (the Start-of-Frame (SOF) and End-of-Frame (EOF) Ordered Sets) are Ordered Sets which immediately precede or follow the contents of a Frame. There are multiple SOF and EOF delimiters defined for the Fabric and N_Port Sequence control.

The two Primitive Signals: Idle and Receiver Ready (R_RDY) are Ordered Sets designated by the standard to have a special meaning. An Idle is a Primitive Signal transmitted on the link to indicate an operational Port facility ready for Frame transmission and reception. The R_RDY Primitive Signal indicates that the interface buffer is available for receiving further Frames.

A Primitive Sequence is an Ordered Set that is transmitted and repeated continuously to indicate specific conditions within a Port or conditions encountered by the receiver logic of a Port. When a Primitive Sequence is received and recognized, a corresponding Primitive Sequence or Idle is transmitted in response. Recognition of a Primitive Sequence requires consecutive detection of 3 instances of the same Ordered Set. The Primitive Sequences supported by the standard are Offline (OLS), Not Operational (NOS), Link Reset (LR) and Link Reset Response (LRR) [1] [2].

5.2 Frame

The basic building blocks of an FC connection are the Frames. The Frames contain the information to be transmitted (Payload), the address of the source and destination ports and link control information. Frames are broadly categorized as Data frames and Link_control frames. Data frames may be used as Link_Data frames and Device_Data frames, link control frames are classified as Acknowledge (ACK) and Link_Response (Busy and Reject) frames. The primary function of the Fabric is, to receive the

Frames from the source port and route them to the destination port. It is the FC-2 layer's responsibility to break the data to be transmitted into Frame size, and reassemble the Frames.

Each Frame begins and ends with a Frame Delimiter (Figure 4) The Frame Header immediately follows the SOF delimiter. The Frame Header is used to control link applications, control device protocol transfers, and detect missing or out of order Frames. An optional header may contain further link control information. A maximum 2112 byte long field (payload) contains the information to be transferred from a source N_Port to a destination N_Port. The 4 bytes Cyclic Redundancy Check (CRC) precedes the EOF delimiter. The CRC is used to detect transmission errors. [1] [2]
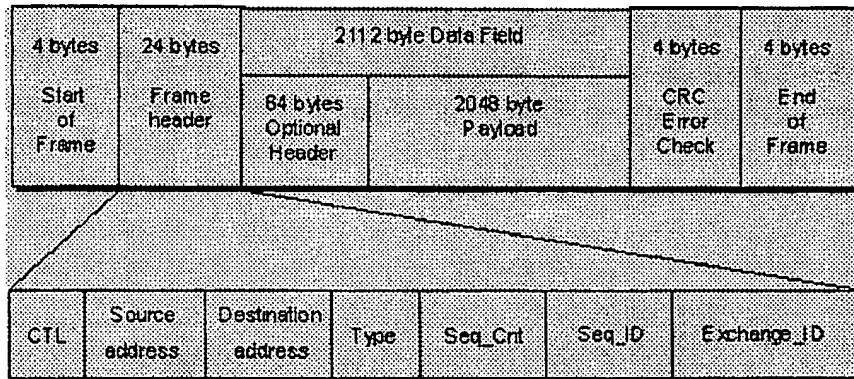
4 bytes	24 bytes	2112 byte Data Field		4 bytes	4 bytes
Start of Frame	Frame header	64 bytes Optional Header	2048 byte Payload	CRC Error Check	End of Frame

CTL	Source address	Destination address	Type	Seq_Cnt	Seq_ID	Exchange_ID

Figure 4 Frame Structure [2]

5.3 Sequence

A Sequence is formed by a set of one or more related Frames transmitted unidirectionally from one N_Port to an other. Each Frame within a sequence is uniquely numbered with a Sequence Count. Error recovery, controlled by an upper protocol layer is usually performed at Sequence boundaries [2].

5.4 Exchange

An Exchange is composed of one or more nonconcurrent sequences for a single operation. The Exchanges may be unidirectional or bidirectional between two N_Ports. Within a single Exchange, only one sequence may be active at any one time, but Sequences of different Exchanges may be concurrently active.

5.5 Protocol

The Protocols are related to the services offered by Fibre Channel. Protocols may be specific to higher-layer services, although Fibre Channel provides its own set of protocols to manage its operating environment for data transfer. The following Protocols are specified by the standard [1]:

- Primitive Sequence Protocols are based on Primitive Sequences (see chapter 5.1) and specified for link failure.
- Fabric Login protocol: The interchanging of Service Parameters of an N_Port with the fabric.
- N_Port Login protocol: Before performing data transfer, the N_Port interchanges its Service Parameters with another N_Port.
- Data transfer protocol describes the methods of transferring Upper Layer Protocol (ULP) data using the Flow control management of Fibre Channel (see chapter 5.6).

- N_Port Logout Protocol is performed when an N_Port requests removal of its Service Parameters from the other N_Port. This may be used to free up resources at the connected N_Port.

5.6 Flow control

Flow control is the FC-2 control process to pace the flow of Frames between N_Ports and between an N_Port and the Fabric to prevent overrun at the receiver. Flow control is dependent upon the service classes (see chapter 5.7). Class 1 Frames use end-to-end flow control, class 3 uses only buffer-to-buffer, class 2 Frames use both types of flow control.

Flow control is managed by the Sequence Initiator (source) and Sequence Recipient (destination) Ports using Credit and Credit_CNT. Credit is the number of buffers allocated to a transmitting Port. The Credit_CNT represents the number of data frames which have not been acknowledged by the Sequence Recipient.

The end-to-end flow control process paces the flow of Frames between N_Ports. In this case the Sequence Recipient is responsible for acknowledging the received valid data Frames by ACK Frames. When the number of receive buffers are insufficient for the incoming Frame, a "Busy", when a Frame with error is received a "Reject" Frame will be sent to the Initiator Port. (see chapter 5.2) The Sequence Initiator is responsible for managing EE_Credit_CNT. The N_Port login (see chapter 5.5) is used to establish EE_Credit.

The buffer-to-buffer flow control is managed between an N_Port and an F_Port or between N_Ports in point-to-point topology. Each port is responsible for managing BB_Credit_CNT. BB_Credit is established during the Fabric Login (see chapter 5.5). The Sequence Recipient (destination) Port signals by sending a Receiver_Ready primitive signal to the transmitting Port whether it has free receive buffers for the incoming Frames.

Figures 5-7 show the flow control management of the different service classes (see chapter 5.7) [1].

5.7 Service Classes

To ensure efficient transmission of different types of traffic, FC defines three classes of service. Users select service classes based on the characteristics of their applications, like packet length and transmission duration, and allocate the services by the Fabric Login protocol.

Class 1 is a service which provides dedicated connections, in effect providing the equivalent of a dedicated physical connection. Once established, a Class 1 connection is retained and guaranteed by the Fabric. This service guarantees the maximum bandwidth between two N_Ports, so this is the best for sustained, high throughput transactions. In Class 1, Frames are delivered to the destination Port in the same order as they are transmitted. Figure 5 shows the flow control management of a Class 1 connection.
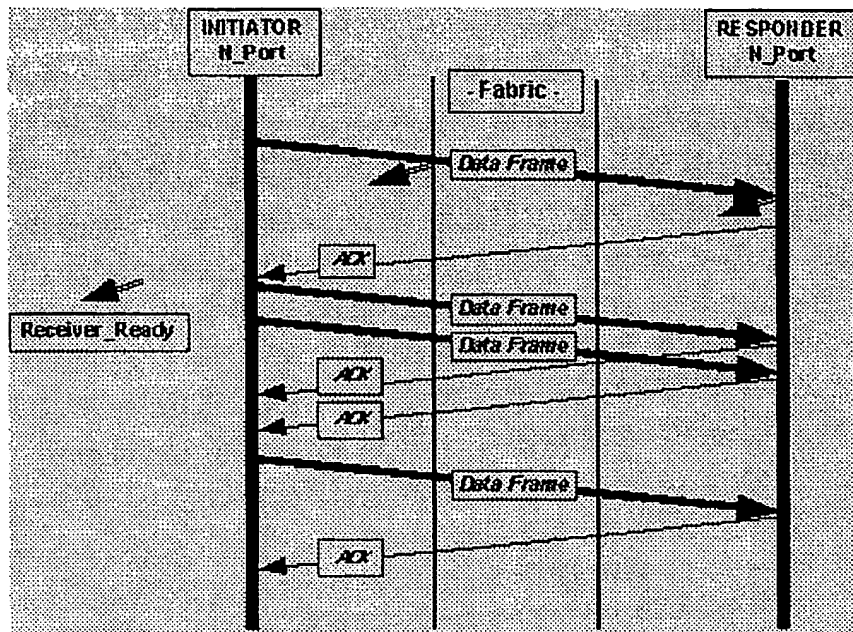
Figure 5 Class 1 Flow Control

Class 2 is a Frame-switched, connectionless service that allows bandwidth to be shared by multiplexing Frames from multiple sources onto the same channel or channels. The Fabric may not guarantee the order of the delivery and Frames may be delivered out of order. This service class can be used, when the connection setup time is greater than the latency of a short message. Both Class 1 and Class 2 send acknowledgment Frames confirming Frame delivery. If delivery cannot be made due to congestion, a Busy frame (see chapter 5.2) is returned and the sender tries again. (Figure 6)
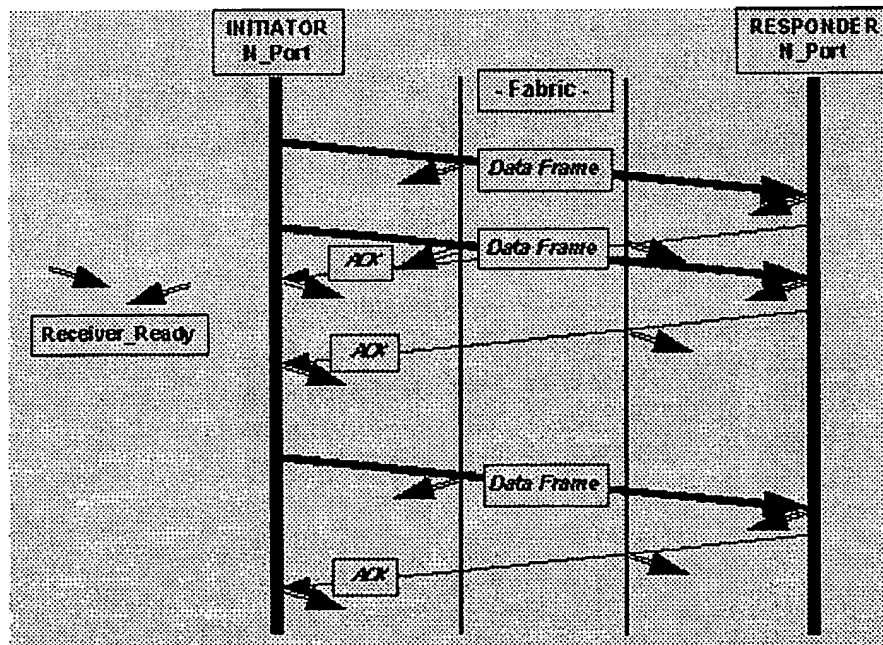

Figure 6 Class 2 Flow Control

Class 3 service is identical to Class 2, except that the Frame delivery is not confirmed. (Flow control is managed only on buffer level, see Figure 7) This type of transfer, known as datagram provides the quickest transmission by not sending confirmation. This service is useful for real- time broadcasts, where timeliness is key and information not received in time is valueless.

The FC standard also defines an optional service mode called intermix. Intermix is an option of Class 1 service, in which Class 1 Frames are guaranteed a special amount of bandwidth, but Class 2 and Class 3 Frames are multiplexed onto the channel, only when sufficient bandwidth is available to share the link [2] [3].
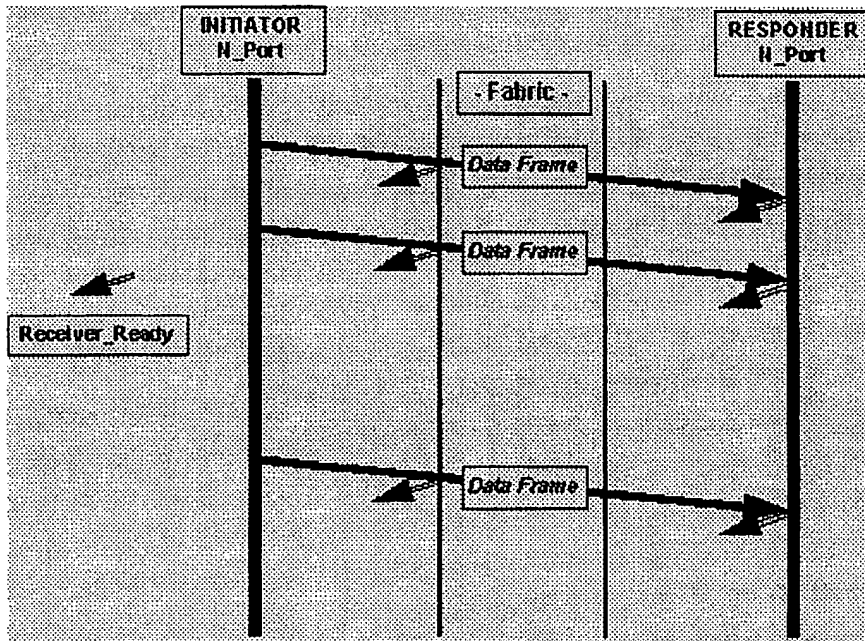

Figure 7 Class 3 Flow Control

6 FC-3 Layer

The FC-3 level of the FC standard is intended to provide the common services required for advanced features such as:

- Striping -To multiply bandwidth using multiple N_ports in parallel to transmit a single information unit across multiple links.
- Hunt groups - The ability for more than one Port to respond to the same alias address. This improves efficiency by decreasing the chance of reaching a busy N_Port.
- Multicast - Multicast delivers a single transmission to multiple destination ports. This includes sending to all N_Ports on a Fabric (broadcast) or to only a subset of the N_Ports on a Fabric. [1]

7 FC-4 Layer

FC-4, the highest level in the FC structure defines the application interfaces that can execute over Fibre

Channel. It specifies the mapping rules of upper layer protocols using the FC levels below. Fibre Channel is equally adept at transporting both network and channel information and allows both protocol types to be concurrently transported over the same physical interface.

The following network and channel protocols are currently specified or proposed as FC-4s [2]:

- Small Computer System Interface (SCSI)
- Intelligent Peripheral Interface (IPI)
- High Performance Parallel Interface (HIPPI) Framing Protocol
- Internet Protocol (IP)
- ATM Adaptation Layer for computer data (AAL5)
- Link Encapsulation (FC-LE)
- Single Byte Command Code Set Mapping (SBCCS)
- IEEE 802.2

Back to Table of Contents

References

1. X3T9.3 Task Group of ANSI: Fibre Channel Physical and Signaling Interface (FC-PH), Rev. 4.2 October 8, 1993
2. Fibre Channel Association: Fibre Channel: Connection to the Future, 1994, ISBN 1-878707- 19-1
3. Gary Kessler: Changing channels, LAN Magazine, December 1993, p69-78

This is one of the CERN High Speed Interconnect pages - 15 August 1994 - Erik van der Bij